International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol.-7, Issue-3, Mar 2019 E-ISSN: 2347-2693

Performance Analysis and Evaluation of Face Recognition using Deep Convolutional Neural Network

Kashyap Patel[1*], Miren Karamta[2], M. B. Potdar[3]

[1]Dept. of Information Technology, Chandubhai S. Patel Institute of Technology, Charusat, Changa, India
[2,3]Bhaskaracharya Institute of Space Application and Geo-Infomatics, Gandhinagar, India

Corresponding author: ksp581994@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i3.10231031 | Available online at: www.ijcseonline.org

Accepted: 12/Mar/2019, Published: 31/Mar/2019

Abstract- Face Recognition system identifies a person by comparing numerous face images accumulated in database records. Face Recognition is simply matching human beings by their faces. This technology has augmented in the field of security and law enforcement to track down criminals and terrorists. In our method, we use deep convolution neural network (deep CNN) and Euclidean distance for extracting the feature from face images. Euclidean Distance used for counting distance between images. We have used FEI dataset for face recognition. This paper gives brief information about face recognition techniques like OpenFace, EigenFace, LBPH, Fisher-Face, and Deep CNN. This paper contains basic information about CNN architecture like AlexNet, GoogleNet, VGGNet, ResNet, SENet, etc. that are used to recognize any type of pose variation in the image. CNN architecture plays an important role to achieve the best accuracy. This paper also focuses on some publicly available datasets: CelebFace (2014), Facebook (2014), Google (2015), MegaFace (2016), MS-Celeb-1M (2018).

Keywords**:** Face Recognition, Face detection, Tensorflow, CNN architectures, Datasets, deep CNN.

I. INTRODUCTION

Face Recognition has become very interesting yet, challenging in the computer vision field due to its wide application possibilities [1]. Using face recognition, it is used to verify, identify and recognize the people which is the main motivation of this system. For the past few years, face recognition has been a stimulating area in the field of security and biometric techniques. There are many algorithms developed in the past few decades. Several feature extraction methods were developed for improving the efficiency of face recognition. In face recognition, many numbers of an evolutionary algorithm like neural networks, VGG16, Face net, and others are used to increase the recognition accuracy. With these algorithms, the application can take data from a given face and can instantly identify the target's face. Face recognition has provided so many opportunities to be increased for research. For example, in this paper, we focus on different types of large face datasets like LFW, YTF, Celeb Faces, and others, which is publicly available on the web and designed perfectly for face identification and verification. In addition, this paper investigates various CNN architectures for face recognition. Nowadays, many recent works on face recognition proposed numerous variants of CNN architecture for a face and we assess some of these modeling choices to filter what is important from irrelevant details [2].
The paper structure is as follows. In section I, we are giving a brief introduction of the general face recognition system. Section II presents different face recognition techniques. In section III, we discussed various CNN architecture with their properties. In section IV, the different datasets of face recognition are discussed. In section V, implementation and results, and in section VI, a conclusion and future work are presented.

II. FACE RECOGNITION TECHNIQUES

This section includes basic information about human face recognition techniques that mostly use for face recognition. The methods include are OpenFace, Deep CNN, Neural Network, Eigan face, and FisherFace. Comparisons of FR Methods/technique as shown in table 1.

A. OpenFace:
OpenFace – a first fully open source real-time facial behavior analysis system [3]. OpenFace is a face identifier implementation using Tensorflow. OpenFace is implemented through torch and python for face identification, which also uses a deep neural network and it mainly focuses on FaceNet. In the working of OpenFace, it has a pre-trained model from OpenCV or Dilb for face detection than after the second step transforms the face for the neural network. For representing the face, OpenFace is used the deep neural network on the 128-bit dimensional unit hypersphere. At last, apply a clustering or classification method to the complete face identification task.

B. Deep CNN:
A deep convolution neural network is one type of artificial neural network, which has more than one hidden layer. The deep CNN architecture involves different types of layers and it includes convolution, polling, and fully connected layers. Deep architectures, and CNN's consist of a large number of neurons and multiple levels of latent calculations of non-linearity [6]. Backpropagation (BP) and Stochastic Gradient Descent are two methods, which are trained by CNN. Both the methods are found weights and biases that minimize certain loss functions in order to map the arbitrary inputs to the targeted outputs as closely as possible.

Table 1: Comparison of FR methods/techniques [17]

Name:	**Authors:**	**Main Idea:**	**Problems:**	**Rate of Recognition**
Deep CNN	ALEX Krizhevsky, Ilya sutskever, Geoffrey E. Hintog 2012	-One or more hidden layers. -Nonlinearity -Weight, bias update	- Computer aided detection(CADe) Problem	97.05%
Neural Network	Stak Lawrence, Lee Giles, Ah Chung Tsoi, Andrew D. Back 1997	-Non linearity -400 images of 40 individuals. -The training time is as long as 4 hours	- When the number of person increase. - Single model images recognition test.	96.2%
EigenFace	L. Sirovich and M. Kirby 1987/1990	-used principal component analysis - database contained 2,500 images of 16 Individuals	- different lighting conditions	96%,85 %, 64% Different lighting condition
FisherFace	Belhumear, V.Hespanda, J.Kiregeman D 1997,Turk, M.A.Pentland AP 1991.	-using PCA and Fisher's linear discriminant analysis	- Different illumination - Head Pose	96%
LBPH (Low Resolution)	Ahonen, Hadid R Pietik,2004. Maenpaa, pietikainen, ojala 2000, wagner 2011.	-Center pixel in the center is compared to its neghobers. -Datasets, LR500	-the problem of unconstrained (gender classification)	94%(45px) 90%(35px)
OpenFace	Brandon Amos, Bartosz Ludwiczak, Mahadev Satganaragah	-FR open source library -Based on google's FaceNet system -Pre-trained models from opencv or dilb.	- Forward pass that take substantially more time than the PCA and LBPH	92.9%

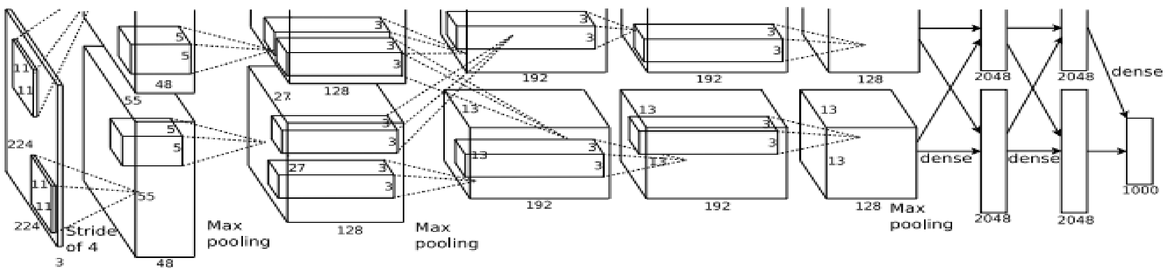
C. Neural Network:
We adopt the working of artificial neural networks by getting inspire by the working mechanism of human brain neurons. The information processing system is a key element of the neural network. For solve a specific problem, a large number of highly connected processing elements are working. The most eye-catching feature of the neural network is its non-linearity. The ORL database is used for correct recognition that contains 400 images of 40 individuals. The training time is 4 hours of features in every layer of the neural network. However, neural network approaches encounter problems when the number of person increase. In addition, for a single model image recognition test but neural networks are not suitable.

D. FisherFace:
Belhumeur et al [7] propose FisherFaces method by using PCA and Fisher's linear discriminant analysis to produce a subspace projection matrix that is very similar to that of the eigenspace method. However, this method can solve one of the main issues that arise in Pentland's EigenFaces method [8]. The leading deficiency of Eigenface methodology is that maximization of scattering is also caused by writing class scatter which can augment detection error-rate only if there is an important fluctuation in the pose on lighting condition

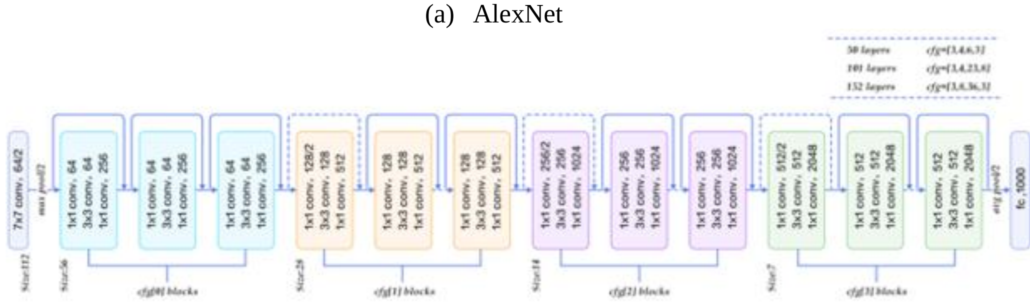
with the same face images. Taking into consideration all these kinds of changes, which occurs are a result of different illumination as well as head pose, which is recorded to be always greater than the variation because of the differences occurring in the face identity procedure, hence it is almost a necessity to develop a robust detection system that can handle every minor deviation. The FisherFaces approach maximizes the class separation but minimizes the variation within classes, which results in overcoming the lightning conditions due to its advantage as mentioned in the first half.

E. EigenFace:
EigenFaces [4] is the technique to a set of eigenvectors and they are used in the human face recognition. This technique developed by Sirovich and Kirby in the year of 1987 for recognition. Eigen picture, eigenvector, and principal component are the key elements for face recognition and use for extracting the feature from the input training images. For efficient represent pictures of faces, Sirovich and Kirby both used principal component analysis (PCA) method. Their database includes 2,500 images of 16 individuals. The author gives results 96%, 85% and 64% correct classifications averaged over lighting. The major problem in this technique is different lighting conditions affect the face recognition process.

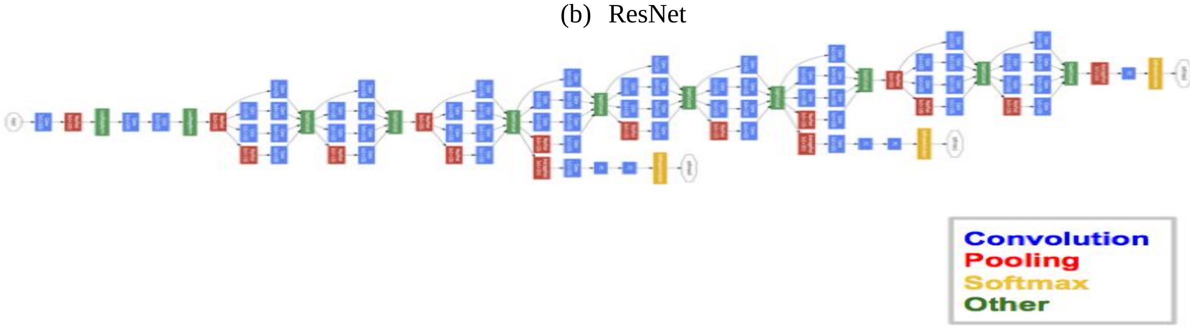
F. Local Binary Patterns Histograms (LBPH):
Local Binary Patterns Histograms [5] is the algorithm architecture, which is used for human face recognition at the low level of resolution. This system better at the minimum low resolution at 35px and the 45px images to identify the various sides poses, angles and other expressions for a human face. This algorithm gives better accuracy on low-resolution images, where a 90% accuracy rate on 35px images and a 94% accuracy rate on 45px images.



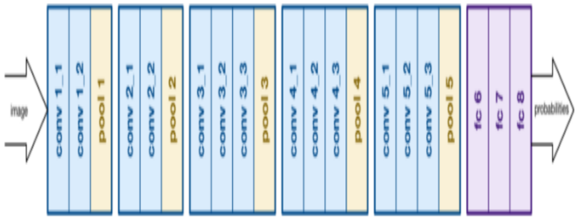
(a) AlexNet



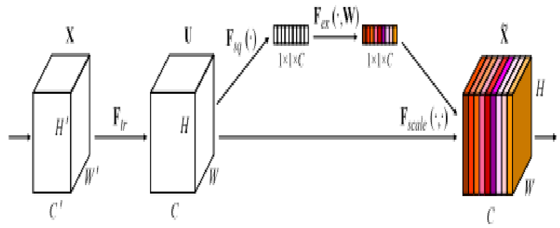
(b) ResNet



(c) GoogleNet



(d) VGGNet



(e) SENet

Figure 1: The CNN architecture of AlexNet, GoogleNet, ResNet, VGGNet, SENet [11].

III. CNN ARCHITECTURE

A Convolutional Neural Network is a special kind of Artificial Neural Network that uses various machine learning algorithms, single and multi-layer perceptron, designed to analyze data. We are discussing different types of CNN architectures: 1) AlexNet, 2) VGG Net, 3) GoogleNet, 4) ResNet, 5) SENet. We present this CNN architecture in fig. 2and also include a comparison of various CNN Architecture in table 2.

A. AlexNet:
AlexNet is a first deeper convolutional neural network, developed in 2012. AlexNet contained eight layers, from those eight layers, the first five layers are convolution layer and the last three layers are fully connected layers. For improve training performance, it used the ReLU activation function. This architecture is used more filters per layer like 11x11, 5x5, 3x3 for deeper network and better performance.

B. VGGNet:
VGGNet is a deep convolution neural network [9], developed by Simonyan and Zisserman in the year of 2014. This network is mainly used in object detection and it achieved better performance in the ImageNet dataset. VGGNet contained 16 Convolution layers and it used 3x3 more filters that extract more features from images. It is publicly available and uniform architecture.
C. GoogleNet:
GoogleNet won the ILSVRC 2014 competition from Google. GoogleNet has 22 layers in its network. From 60 million to 4 million, it reduces the number of parameters. These layers also known as Inception layers. Three number of filter contain by inception layer: 5×5, 3×3 and 1×1. Using this filter, it has improved the performance on deep learning to use more data and more layers. GoogleNet occupies nine inception layers in the CNN network. The normalization, image distortions, and RMSprop are the major areas where GoogleNet is used.

D. ResNet:
ResNet stands for Residual Neural Network. ResNet won the ILSVRC 2015 classification competition with a top-5 error rate of 3.57%. This architecture was learned with network depth of as large as 152. This architecture uses 3x3 filter same as VGGNet. It gave better accuracy performance than VGGNet and GoogleNet.

E. SENet:
SENet stands for the Squeeze-and-Excitation Neural Network Model. SENets won first place the ILSVRC 2017 competition, which significantly reduced the top-5 error to 2.251%, achieving a 25% relative improvement over the winning entry of 2016 [10]. Squeeze-and-Excitation Networks introduces a building block for CNN, which can improve channel interdependencies at almost no computational cost.

Table 2: The comparison of various CNN Architecture

Year	CNN Architecture	Top 5 error rate	Number of Parameter	CNN layer
2012	AlexNet	15.3%	60 Million	8 layers
2014	VGG Net	7.3%	138 Million	19 layers
2014	Google Net	6.66%	4 Million	22 layers
2015	ResNet	3.57%	25 Million	152 layer
2017	SENet	2.25%	Above 25 Million	12 layer

IV. DATABASE

In the past decades, many face databases have been constructed with a clear tendency from small-scale to large-scale, from the single source to diverse-sources, and from controlled lab to real-world unconstrained condition [9]. There are several types of face dataset built from the past few years for face verification, face identification, and face recognition. For example, VGG face, MegaFace, Google dataset, MS-Celeb-1M, Facebook [15] dataset [12] and so on.

In MS-Celeb-1M Dataset, this is one of the largest publicly available dataset, developed for challenge 1 & 2 in 2016 and for challenge 3 in 2018. They design a benchmark task to recognize one million celebrities from their face images and identify them by linking to the unique entity keys in a knowledge base [13], which contains ten million images. VGG Face and VGG Face 2 are developed in 2015 and in 2017, which is publicly present on the web. 2622 unique photos and 2.6 million face photos are in the VGG Face dataset and 9131 different images and 3.31 million face images in VGG Face 2 [16] include. Presently, MegaFace is one of the most difficult datasets for face recognition because of almost 1 million distractors and 690,572 unique users present in their dataset. MegaFace includes 672,057 unique identities and 4.7 million photos from their unique identities. MegaFace are selected photos from the test set: face scrub Celebrities and FGNet Age-invariant non-celebrities.

Google and Facebook both are privately available datasets. Google is the largest dataset with 8 million people photos, and 200M faces images and Facebook has contained 4k person images and the size of the dataset is 4400k. Both datasets can major impact on the face recognition system. The CASIA-WebFace is currently the largest dataset, which is publicly available, with about 10K celebrities, and 500K images [13].

Table 3: datasets of face images for face recognition

Year	**Dataset Name**	**Available: (Public or Private)**	**Unique Identities**	**Face Images**
2014	CelebFace	Public	10K	202K
2014	Casia-WebFace	Public	10K	500K
2014	Facebook Dataset	Private	4K	4400K
2015	VGG Face	Public	2622	2.6M
2015	Google Dataset	Private	8M	200M
2016	MegaFace	Public	672,057	4.7M
2017	VGG Face 2	Public	9131	3.31M
2018	MS-Celeb-1M (for Challenge 3)	Public	100K	10M

V. IMPLEMENTATION AND RESULTS

In this research, we try to resolve the problem of face recognition by using CNN architectures. CNN architecture uses different techniques to recognize the face. FR system collects all the feature from the given input image and matches input face image feature with labeled face in the database. For extracting face image feature, we use deep convolution neural network (deep CNN) and Euclidean distance. Euclidean distance mainly used for map distance between training images and testing images. In this implementation, we try to use OpenCV, Facenet and Dlib techniques for detect, image transformation and measure the distance between training image and testing image. We used python Keras and Tensorflow environments for developing the FR system model.

Here, the aim is to train our model by giving training images to detect the face and to identify the distance between the training image and the testing image. Here we use the threshold value = 0.56 to differentiate images using calculated distance value. If the calculated distance between the training image and the testing image is high than the threshold both are a different person and if the calculated distance between the training image and the testing image is low then the threshold both are the same person. Here distance value can depend on triplet loss L function, where loss function is used for counting distance and also for comparing a training image and testing image. Triplet loss function is the following:

$$L = \sum_{i=1}^{m} \left[\left\| f(x_i^a) - f(x_i^p) \right\|_2^2 - \left\| f(x_i^a) - f(x_i^n) \right\|_2^2 + \alpha \right]_+ \quad (1)$$

Here, m is the number of triplet in the training set. This training set is most important for selecting positive pairs (same person image) and negative pairs (different person image). This calculated distance mainly depends on the threshold value. We obtained 128-dimensional embedding vectors from CNN architecture's two top layers which are referred as embedding layers. In the pre-trained network, embedding vectors can be calculated by feeding the aligned and scaled images. Using training and testing images we can calculate the distance threshold values of them. An optimal threshold value is the one constant value, which can be used to decide low or high distance between the input image and output image. So face recognition and face verification accuracy are mainly depend on a threshold value.

Here as shown in figure 2, we detect the face area of an image. We put a bounding box on the area of face after detection. As shown in figure 3, after detecting a face we try to identify or recognize the person or identity by calculating the distance between the images. If distance lies between the threshold range than they are the same identity, otherwise different identity. In figure 4, we show Graphical Representation of Face Recognition Results (Threshold and Accuracy) and In figure 5, we gave Graphical Representation of Face Recognition Accuracy and Distance which is measures between two same face images.

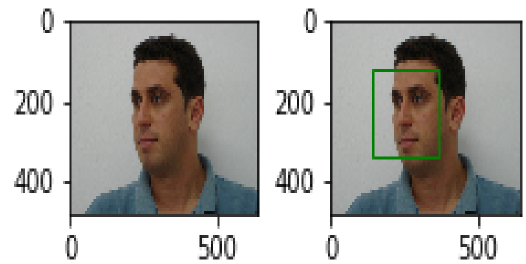


Figure 2: Face Detection

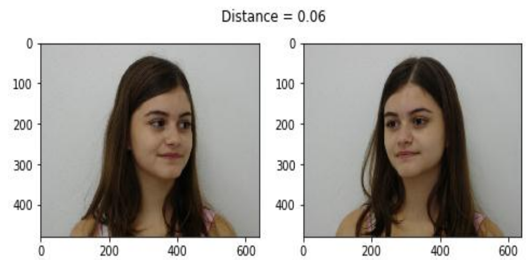


Image 1

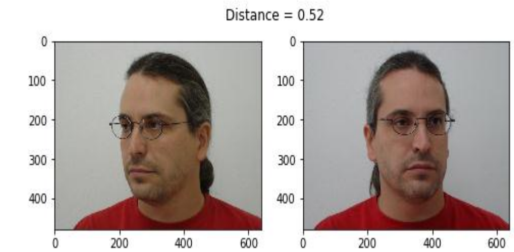


Image 2

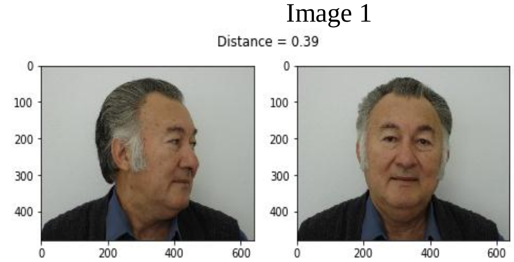


Image 3

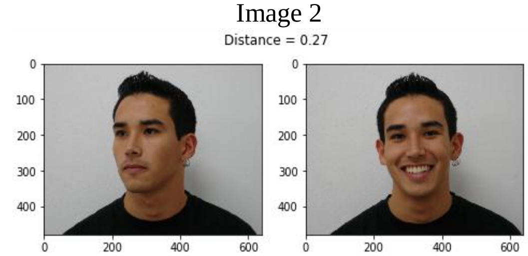


Image 4

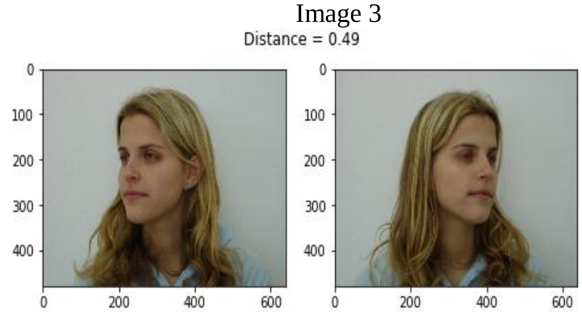


Image 5

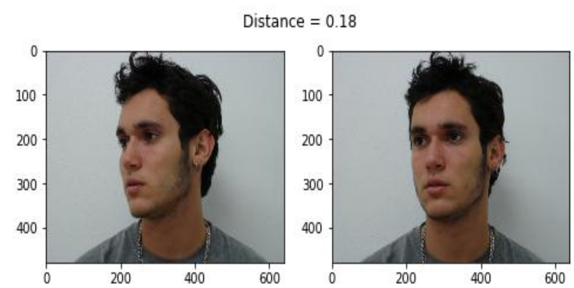


Image 6

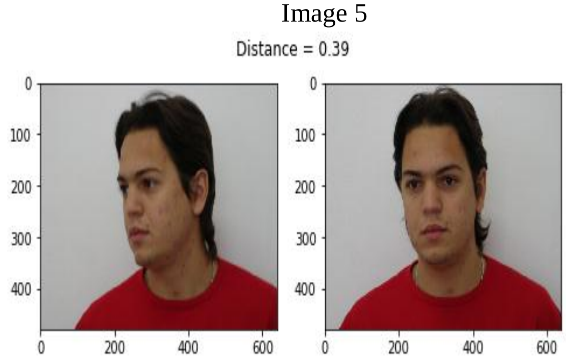


Image 7

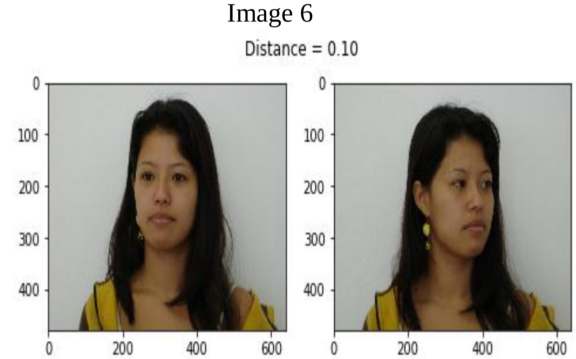


Image 8

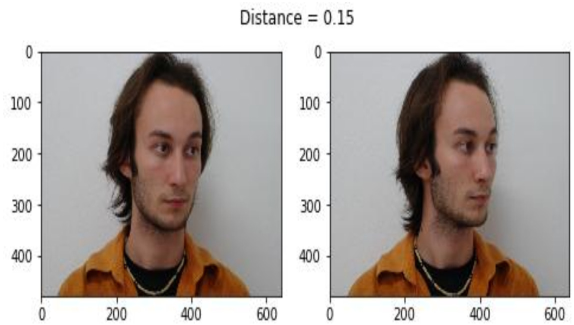


Image 9

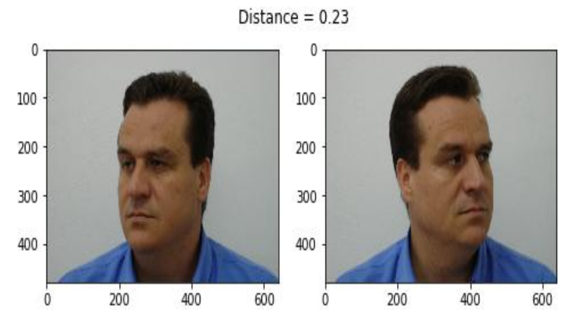


Image 10

Figure 3: Result of Face Recognition

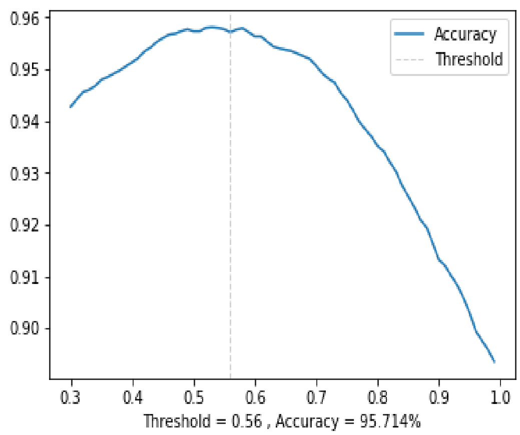


Image 1

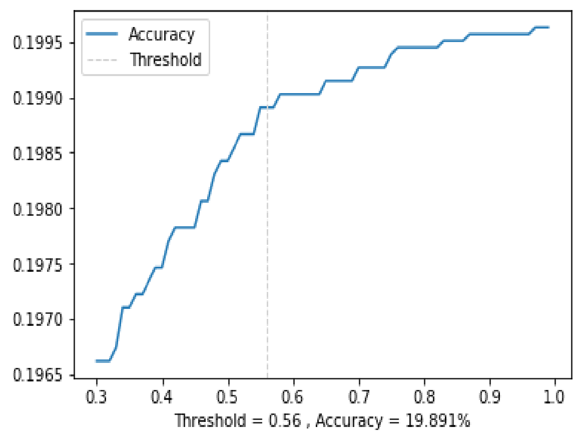


Image 2

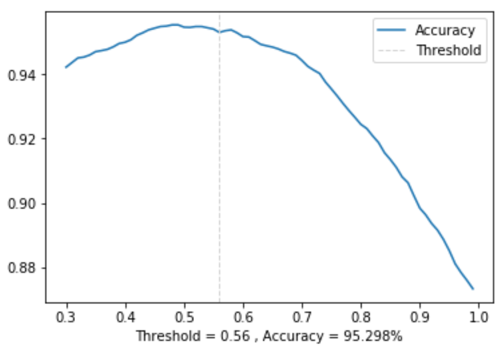


Image 3

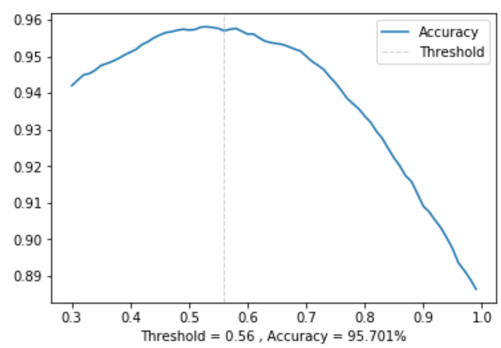
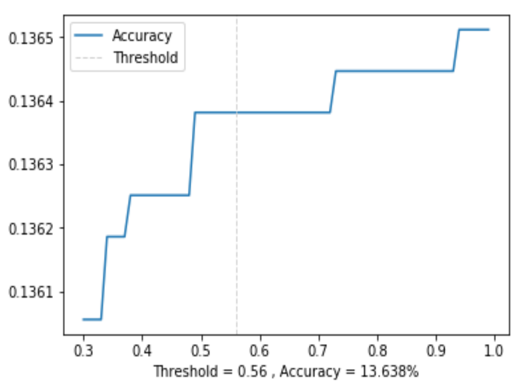


Image 4



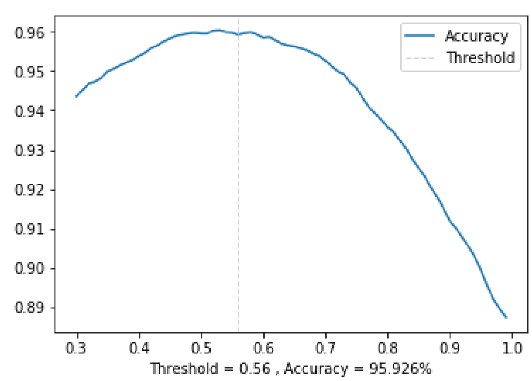


Image 5

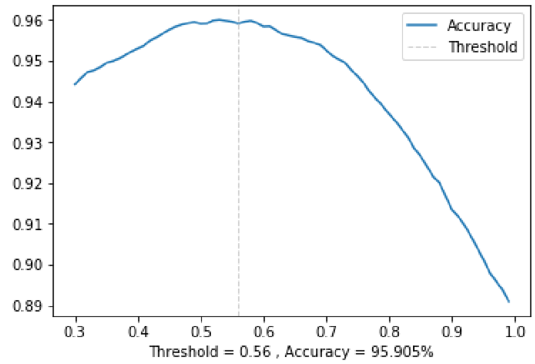


Image 6

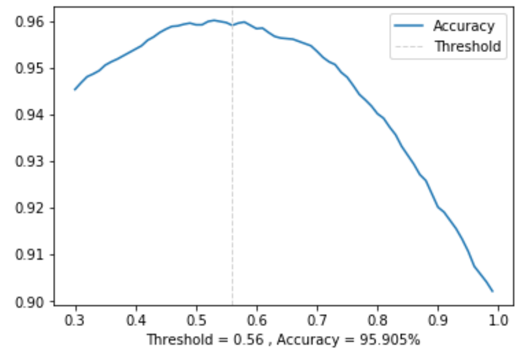
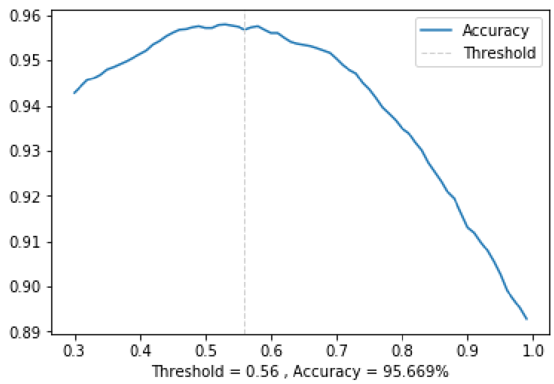


Image 7

Image 8



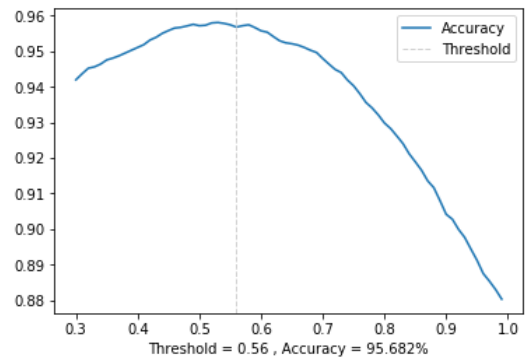


Image 9

Image 10

Figure 4: Graphical Representation of Face Recognition Results (Threshold and Accuracy).

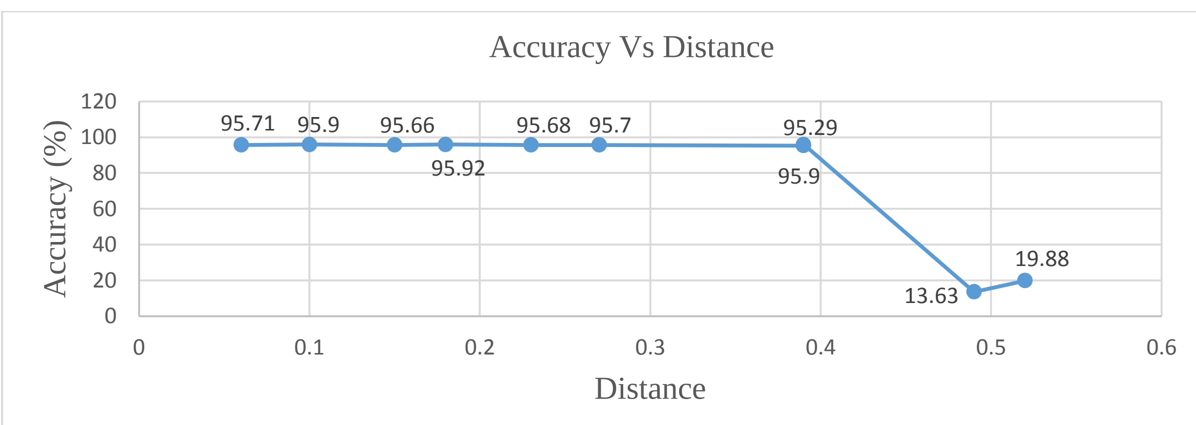


Figure 5: Graphical Representation of Face Recognition Accuracy and Distance which is measures between two same face images.

Table 4: Results of two same matching images

Images	**The Distance between two images**	**Threshold value**	**Accuracy (%)**
Image 1	0.06	0.56	95.714
Image 2	0.52	0.56	19.881
Image 3	0.39	0.56	95.298
Image 4	0.27	0.56	95.701
Image 5	0.49	0.56	13.638
Image 6	0.18	0.56	95.926

Image 7	0.39	0.56	95.905
Image 8	0.10	0.56	95.905
Image 9	0.15	0.56	95.669
Image 10	0.23	0.56	95.682

VI. CONCLUSION AND FUTURE WORKS

In this paper, some of the effective face recognition techniques were presented. The comparative analysis of various CNN architectures with the detail of the top five error rate, parameters and applications show that this technique and architecture both are used to increase the accuracy and recognition rate of a test image. It can work with any kind of images and is reasonably robust to changes in facial expression or orientation, light conditions, and other factors.

In the future, we will extensively test this technique using an image dataset with different parameters and combinations by using a pose-awareness [14] method, then after using a large dataset; it will achieve comparatively best accuracy using less amount of time. We will also try to recognize people in videos in which after processing, it draws a bounding box and identifies the person. The system will count the number of people using their face in a particular area. There is also a scope of integrating GIS into the face recognition system. So there is still a lot of work to do with this project but our main focus is to develop a face recognition system, which can achieve all three goals.

REFERENCES

[1]. Anil, J., and L. Padma Suresh. "Literature survey on face and face expression recognition." Circuit, Power and Computing Technologies (ICCPCT), 2016 International Conference on. IEEE, 2016.

[2]. Omkar M. Parkhi, Andrea Vedaldi and Andrew Zisserman, "Deep Face Recognition," in British Machine Vision Conference (BMVC), 2015.

[3]. Tadas Baltruˇsaitis, Peter Robinson and Louis-Philippe Morency, "OpenFace: an open source facial behavior analysis toolkit," in 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 2016

[4]. Sawant, Mr. Namdev M., Mr. Rirandra A. Dhotre, and Mr. Pravin P. Botre. "REVIEW ON PARALLEL IMPLEMENTATION OF PRINCIPAL COMPONENT ANALYSIS FOR FACE RECOGNITION USING MPIAND OpenMP." GSJ 6.8 (2018): 95

[5]. Ahmed, Aftab, et al. "LBPH based improved face recognition at low resolution." 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD). IEEE, 2018

[6]. Ivars Namatēvs, "Deep Convolutional Neural Networks: Structure, Feature Extraction, and Training," 2017

[7]. Belhumeur, V., Hespanda, J., Kiregeman, D.,1997, Eigenfaces vs. FisherFaces: recognition using class specific linear projection, IEEE Trans. on PAMI, V. 19, pp. 711-720.

[8]. Turk, M. A., Pentland, A. P., Eigenfaces for recognition, 1991, Cognitive Neurosci., V. 3, no.1, pp.71-86

[9]. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[10]. Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." arXiv preprint arXiv:1709.01507 7 (2017).

[11]. H. Shin *et al.*, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," in *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285-1298, May 2016.

[12]. Wang, Mei, and Weihong Deng. "Deep Face Recognition: A Survey." arXiv preprint arXiv:1804.06655 (2018).

[13]. Guo, Yandong, et al. "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition." European Conference on Computer Vision. Springer, Cham, 2016.

[14]. Masi, Iacopo, et al. "Learning pose-aware models for pose-invariant face recognition in the wild." IEEE Transactions on Pattern Analysis and Machine Intelligence (2018).

[15]. Hmani, Mohamed Amine, and Dijana Petrovska-Delacrétaz. "State-of-the-art face recognition performance using publicly available software and datasets." Advanced Technologies for Signal and Image Processing (ATSIP), 2018 4th International Conference on. IEEE, 2018.

[16]. Cao, Qiong, et al. "Vggface2: A dataset for recognizing faces across pose and age." Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on. IEEE, 2018.

[17]. Bagherian, Elham, and Rahmita Wirza OK Rahmat. "Facial feature extraction for face recognition: a review." Information Technology, 2008. ITSim 2008. International Symposium on. Vol. 2.IEEE, 2008.

Authors Profile

Mr. Kashyap Patel completed his bachlor degree in Information Technology from Charusat University, Changa, Gujarat, India in the year of 2016. Now pursuing master degree in Information Technology under the guidance of Mr. Miren Karamta Project Scientist, Bhaskaracharya Institute of Space Application and Geo-Infomatics, Gandhinagar, India

Mr. Miren Karamta is working as a Project Scientist and IT Systems Manager at the Bhaskaracharya Institute for Space Applications and Geo-informatics (BISAG), Gandhinagar, Gujarat since January 2011. He is a 2010 MTech in Computer Engineering from DDIT.

Dr. M. B. Potdar is a 1982 Ph. D. in Physics from Physical Research Laboratory of Dept. of Space, Govt. of India. Later for 28 years, he was associated with the Indian Space Research Organization in various capacities. He worked extensively in development of land and atmospheric applications of Remote Sensing data. Since March 2011, he is holding position as Project Director at BISAG and organizing and steering research in various areas of software development and applications of geo-spatial data.

